

AUDIO FEATURES DESCRIPTION METHOD AND AUDIO VIDEO FEATURES
DESCRIPTION COLLECTION CONSTRUCTION METHOD

BACKGROUND OF THE INVENTION

5 Field of the Invention

The present invention relates to a method of describing the features of compressed or uncompressed audio data and a method of constructing the feature description collection of compressed or uncompressed audio video data. The audio
10 feature description method is a method of describe audio feature attached to audio data and enables high-speed, efficiently search and browse audio data at various levels from coarse levels to fine levels. Also, the audio video feature description collection construction method is a
15 method of collecting the feature descriptions of multiple items of audio video data according to a specific feature type, and constructing multiple feature descriptions conforming to the specific feature type as a feature description collection, thereby making it possible to
20 acquire a feature description collection based on the specific feature type from multiple audio video programs.

Description of the Related Art

The description of the features of audio data can represent the features of the entire audio data with a small
25 quantity of feature by describing or combining the spatial features or frequency features of a audio file existing as a compressed or uncompressed file. The feature description

can be employed as an element for grasp the summary of audio data when searching the audio data. The feature description is effective when searching desired audio data from as an audio database and browsing the content of the desired audio data.

Conventionally, methods of describing features have been considered mainly for video information. The considerations have been, however, only given to how to represent feature values for audio data. How to combine which feature values so as to describe entire audio data has not been specified or considered yet.

Meanwhile, the description of the features of audio video data has been currently studied at MPEG-7 (Motion Picture Coding Experts Group Phase 7) in ISO (International Organization for Standard). In the MPEG-7, the standardization of content descriptions and description definition languages for allowing efficient search to compressed or uncompressed audio video data is now underway.

In the MPEG-7, feature descriptions from various viewpoints are standardized. Among the feature descriptions, summary description allowing high-speed, efficient browsing of audio video data is allowed to describe only information for a single audio video in the MPEG-7. Due to this, summary information according to various summary types on a single audio video program can be constructed and described. Summary types involve important events of the program, important audio clips, video clips and so on.

For example, as shown in Fig. 22A and 22B, for single audio video programs 50 and 51, i.e., complete audio video programs 50 and 51, summary information on various summary types, e.g., "home run", "scoring scene", "base stealing scene" and "strike-out scene", can be described as a summary collection.

As for summary description, for example, among conventional features descriptions of audio video data, summary information only for a single video audio program can be constructed and described as shown above. However, the construction and description of summary information for multiple audio video programs are not currently specified.

Further, if a feature description collection is described using the feature descriptions of a summary collection from multiple programs in a currently specified framework, e.g., if a feature description collection is described using the feature descriptions of a summary collection from, for example, multiple programs 50, 51, as shown in Fig. 22A or 22B, then the feature description collection is expected to be described as shown in, for example, Fig. 15A. Namely, it is expected that summary information on the summary collection for each program are simply collected and described.

Consequently, the conventional feature description collection tends to be redundant and unnecessary processings are carried out to search a desired summary from the summary collection, making disadvantageously search time longer.

Further, it is difficult to clearly describe the designations of programs to be referred to for each summary. Besides, in case of searching a desired summary from the summary collection, it is difficult to represent a combination of
5 multiple summary types.

SUMMARY OF THE INVENTION

It is, therefore, an object of the present invention to provide a feature description method capable of high-speed,
10 efficiently searching audio data or grasping the summary thereof by giving consideration to elements and features specific to audio data. It is another object of the present invention to provide a method of constructing an audio video feature description collection for collecting the feature
15 descriptions for multiple audio video programs according to a specific feature type to thereby make it possible to efficiently, clearly describe a feature description collection. It is yet another object of the present invention to provide a method of constructing an audio video
20 feature description collection capable of acquiring a desired feature description from a feature description collection by combining multiple feature types.

In order to achieve the above object, the first feature of the present invention is that audio features are
25 hierarchically represented by setting an audio program which means entire audio data constructing one audio program as a highest hierarchy and describing the audio features in

a order from higher to lower hierarchies, said hierarchies being represented by at least one audio program having a semantically continuous content and at least one of an audio scene and an audio shot, and said hierarchies being described
5 by at least names of the hierarchies, audio data types, feature types and feature values described by audio segment information classified according to the feature types.

According to these features, compressed or uncompressed audio data can be described hierarchically by using novel
10 method. Besides, it is possible to provide compressed or uncompressed audio feature description capable of high-speed, efficiently searching or inspecting audio data.

The second feature of the invention is that a compressed or uncompressed audio video feature description collection
15 construction method, wherein feature descriptions based on multiple feature types are associated with each audio video program; the feature descriptions are extracted from multiple audio video programs based on a specific feature type; a feature description collection is constructed by
20 using multiple extracted feature descriptions; and the feature description collection is described as a feature description collection file.

And the third feature of the invention is that the feature type is a summary type; summary descriptions associated with
25 the individual audio video programs are extracted from multiple audio video programs based on a specific summary type; a summary collection is constructed using multiple

extracted summary descriptions; and the summary collection is described as a summary collection file.

According to the second and third features, the feature descriptions from multiple audio video programs are collected according to a specific information type, therefore the feature description collection can be represented efficiently and clearly. Further, it is possible to combine multiple feature types and to obtain a desired feature description from the feature description collection.

BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1 is a block diagram showing the functionality of one embodiment according to the present invention;

Fig. 2 shows an example of constructing an audio data (music program) hierarchy;

Fig. 3 is a block diagram showing one example of the internal structure of a feature description section shown in Fig. 1;

Fig. 4 is a block diagram showing one example of the internal structure of an audio element extraction section shown in Fig. 3;

Fig. 5 shows an example of the description format in a scene description section, a shot description section and a clip description section;

Fig. 6 is an illustration showing the example of the format shown in Fig. 5 applied to the structure of Fig. 2;

Fig. 7 shows an example of the format of the key audio clip, the key stream and the key object;

Fig. 8 is an illustration showing the key stream and the key object applied to the structure of Fig. 2;

5 Fig. 9 shows an example of the format of the key event, audio slides and audio thumbnails;

Fig. 10 is a block diagram showing the internal structure of a feature extraction section shown in Fig. 1;

10 Fig. 11 is a block diagram showing an alternative of the present invention;

Fig. 12 shows an example of the format of the key audio clip attached a level structure;

15 Fig. 13 is a block diagram showing the diagrammatic sketch and a processing flow of another embodiment according to the present invention;

Fig. 14 is a conceptual illustration of a summary collection constructed in a feature description collection construction section shown in Fig. 1;

20 Fig. 15 shows an example of the description contents of feature description collection files obtained by the conventional method and by the method of the present invention;

25 Fig. 16 shows an example of the description contents of the feature description files obtained by the conventional method and by the method of the present invention in a table form;

Fig. 17 is a block diagram showing a diagrammatic sketch

and a processing flow if a feature type is a summary type shown in Fig. 13;

Fig. 18 shows an example of the description contents of summary collection files obtained by the conventional
5 method and by the method of the present invention;

Fig. 19 shows another example of the description contents of the feature description collection files obtained by the conventional method and the method of the present invention;

10 Fig. 20 is illustration if "feature type" shown in Fig. 19 is a "summary type";

Fig. 21 is a flow chart showing an operation for generating a nested summary collection file; and

Fig. 22 is illustration for a summary collection
15 generated by the conventional method.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

The present invention will be described hereinafter in detail with reference to the accompanying drawings.
20 First, the definition of terms used in the present invention will be described.

"Audio program (or audio file)" ... the entirety of audio data constructing one audio program.

"Audio segment" ... a group of adjacent audio
25 samples in an audio program.

"Audio scene" ... temporally and semantically continuous audio segments. Group of audio shots.

"Audio shot" ... audio segments which are temporally and semantically continuous to adjacent audio segments but which have different characteristic from that of adjacent audio segment. Characteristics involve an audio data type, a speaker type and so on.

"Audio clip" ... audio segments which are temporally continuous and have one meaning.

"Audio stream" ... each audio data for each channel or track when the audio data consists of multiple channels or tracks.

"Audio object" ... audio data source and subject of auditory event. The audio data source of an audio stream is an audio object.

"Audio event" ... behavior of an audio object in a certain period or an auditory particular event or audio data attached to visual particular event.

"Audio slide" ... audio data consisting of a sequence of audio pieces or audio programs and obtained by playing these audio pieces or audio programs at certain intervals.

The present invention is based on a conception that audio data is represented by a hierarchical structure. An example of the hierarchical structure will be explained referring to Fig. 2.

As shown in Fig. 2, a compressed or uncompressed audio program or audio program file (a) (to be referred to as "audio program (a)" hereinafter) (the first hierarchy) which is, for example, a "music program" can be represented by, for

example, audio scenes (b) (the second hierarchy) consisting of "interview scene 1" and "vocal scene 1". The "interview scene 1" in the audio scenes (b) can be represented by audio shots (c) (the third hierarchy) consisting of "MC's talks", "singer's talks", ..., "plaudits" and also the "vocal scene 1" can be represented by the audio shots (c) (the third hierarchy) consisting of "melody 1", ..., and "melody 4". Also, "topic 1", "topic 2", "introduction" and so on which are distinctive parts extracted from the audio program (a), audio scenes (b) or audio shots (c), can be represented by audio clips (d) (the fourth hierarchy). Further, if the "melody 2", for example, in the audio shots (c) consists of signals of multiple channels or track, the "melody 2" can be represented as audio stream. Each audio stream can be represented as audio objects such as "voice", "piano", "guitar" and so on.

Next, one embodiment of a function which realizes the method of the present invention will be explained referring to Figs. 1 and 2.

In this embodiment, description will be given to a feature description, among feature descriptions of audio data, relating to summary (outline) for high-speed, efficiently grasping the outline of the audio data.

First, if a compressed or uncompressed audio program or audio file (a) (to be referred to as "audio program (a)" hereinafter) is inputted into a feature description section 1 shown in Fig. 1, as shown in Fig. 2, the audio program

(a) is divided into a single or multiple audio scenes (b) which are semantically continuous and the audio scenes (b) are divided and hierarchically structured into a single or multiple audio shots (c). Further, the audio shots are
5 divided into audio clips (d) which have one meaning respectively and described hierarchically. The hierarchies under the audio program (a) are not necessarily essential and are not necessarily ordered as shown above. Thereafter, a feature description file 1a which describes
10 the entire audio program (a), is generated according to a feature type.

These hierarchies are described by at least the names of each hierarchy and/or the feature values thereof. Feature values include feature types, audio data types and
15 audio segment information corresponding to the feature types. The audio segment information is described by any of time codes for start time and end time, time codes for start time and duration, a start audio frame number and an end frame number, or a start frame number and number of frames
20 corresponding to duration. The segmentation of the audio program (a) and structurization into hierarchies can be performed manually or automatically.

Further, the feature description section 1 generates a thumbnail 1b for describing the audio program (a) as either
25 audio pieces or images. The thumbnail 1b consists of a description indicating a thumbnail, and the segments or file names of the audio pieces or the file names of the images.

The audio program (a), feature description file 1a and thumbnail 1b are inputted into a feature extraction section 2. The feature extraction section 2 searches the corresponding portion of the feature description file by search query 2a from a user and performs feature presentation 2b. If the feature type of the search query 2a is the thumbnail 1b, the thumbnail is presented. If the feature type is a type other than the thumbnail, segments described in the feature description file 1a are extracted from the audio program and presented.

Fig. 3 is a block diagram showing the internal structure of the feature description section 1. If the audio program (a) is inputted into the feature description section 1, the audio program (a) is fed to an audio element extraction section 11. The internal structure of the audio element extraction section 11 is shown in Fig. 4. The audio program (a) inputted into the audio element extraction section 11 is divided into scenes in a scene detection section 111 and the those scenes are further divided into shots in a shot detection section 112. Scene information and shot information generated from the scene detection section 111 and the shot detection section 112 include indication of scene or shot, and each segment information.

Further, if audio data consists of multiple channels
25 or tracks, a stream extraction section 113 extracts each
channel or track as a stream and outputs stream information.
Stream information include stream identifiers and segment

00720667, 120700

information for each stream. An object identifying section 114 identifies an object as the audio source of the stream from each audio stream and outputs object information. The objects include, for example, "voice", "piano", "guitar" and so on (see Fig. 2). The object information includes the stream identifier and content of object as well as audio segment information corresponding to the object.

An event extraction section 115 extracts an event representing a certain event from the audio program (a) and generates, as event information, the content of the event and audio segment information corresponding to the event.

A slide extraction section 116 extracts audio pieces which are introductions or representative of the audio program, audio scene or audio shot, and outputs, as slide information, information for each audio piece. The slide information includes segment information if the audio slide components are audio segments, and includes file names if the audio slide components are audio files.

The extraction of each information in the audio element extraction section 11 shown in Figs. 3 and 4 can be also conducted manually.

The information generated from each section in Fig. 4 are inputted into corresponding description sections shown in Fig. 3. First, the scene information and shot information are inputted into a scene description section 12 and a shot description section 13, respectively. The scene description section 12 and the shot description section 13

describe the types of scenes and shots belonging to the audio program (a), the audio data type and its segment information, respectively. A clip extraction section 14 extracts, as a clip, an audio piece having a certain meaning among the
5 scenes or shots. If necessary, a clip description section 15 declares and describes a clip as the feature type, the audio data type and its segment information.

Figs. 5 and 6 shows an example formats of the description in the scene description section 12, the shot description
10 section 13 and the clip description section 15. Fig. 5 shows an example formats represented generally, and Fig. 6 shows an example represented according to the structure of Fig. 2.

As for the clips shown above, a particularly important
15 clip in the program is regarded as the key audio clip. A key clip description section 16 declares and describes a key audio clip as a feature type, as well as the audio data type and its segment information. Fig. 7(a) shows an example format for describing the key audio clips.

20 Further, among the key audio clips, distinctive voice, music and sound are regarded as the keyword, the key note and the key sound, respectively, and a key audio clip is described as a feature type and the audio data type and its segment information are also described. As for the keyword,
25 the content of the speech is simply described as text information. Figs. 7(b), 7(c) and 7(d) show an example format for describing the key word, the key note and the

key sound, respectively. Key words involve, for example, speeches saying such as "year 2000", "Academy Award". Key notes involve, for example, a "main theme" part of music. Key sounds involve, for example, the sound of "plaudits".

5 Meanwhile, stream information and object information are inputted into a stream description section 17 and an object description section 19 shown in Fig. 3, respectively. Among the streams and objects, particularly important stream and object are regarded as the key stream and the key object, 10 respectively. The feature type of the key stream and that of the key object as well as the audio data type, the contents of feature values and segment information are described by a key stream description section 18 and a key object description section 20, respectively. Figs. 7(e) and 7(f) 15 shows an example format for describing the key stream and key object, respectively. Figs. 8(a) and 8(b) shows an example format for describing the key stream and key object according to the structure of Fig. 2. The content of the key object is described by text information.

20 Further, event information is inputted into an event description section 21. A representative event is regarded as the key event. The feature type of the key event, the audio data type, the contents of feature values and segment information are described by a key event description section 25 22. Fig. 9(a) shows an example format for describing the key event. The content of the key event is described by text information. Key events involve, for example,

"explosion" and words like "goal" in soccer game program.

Furthermore, slide information is inputted into a slide construction section 23. The slide construction section 23 constructs an audio slide from multiple audio pieces included in the slide information. The content of the audio slide is described by a slide description section 24. The slide description section 24 describes the type of features, audio segments or the names of files constructing the audio slide. The content of the description relating to the audio slide is also constructed as a feature description file. Figs. 9(b) and 9(c) show an example format for describing the audio slide.

In addition, a thumbnail generation section 25 (see Fig. 3) belonging to the same level as that of a program hierarchy section generates a thumbnail lb representing the content of the audio program from the audio program. The thumbnail lb may be represented by a single or multiple audio pieces or as images. Figs. 9(d) and 9(e) show an example format for describing the audio thumbnail.

As described above, all the description contents generated from each description section shown in Fig. 3 are components of the feature description file 1a.

If the feature type of the audio data is a shot or a key audio clip (including a key word, a key note and a key sound), it is possible to add values indicating hierarchical levels in the same feature type, and to search and browse hierarchically multiple pieces of audio data with the same

feature type according to the level values. As an example of describing levels, level 0 is a coarse level and level 1 is a fine level. It is possible to specify audio segments having corresponding feature types for each level. Level
5 information can be specified, for example, between the audio data type and the audio segments as shown in Figs. 12(a) through 12(d). Moreover, if the audio segment belonging to the level 0 also belongs to the level 1, the description indicating that situation at the same level as that of the
10 feature type makes it possible to avoid overlapping of audio segments. Thus, it is possible to describe multiple levels according to a common feature type and an audio data type, and to specify audio segments according to level values.

Fig. 10 is a block diagram showing the internal
15 structure of a feature extraction section 2 (see Fig. 1). The audio program (a), the feature description file 1a outputted from the feature description section 1, the thumbnail 1b and the search query 2a as input information from the user are inputted into the feature extraction
20 section 2. First, the feature description file 1a is loaded into a feature description file parsing section 41 which parses a feature type, an audio data type, its segment information and so on.

Next, based on the search query 2a inputted from the
25 user and information from the feature description file parsing section 41, a feature description matching section 42 searches the feature specified by user and outputs the

specified segments of the audio program (a) described as a corresponding feature type.

A feature extraction section 43 extracts audio data according to actual feature values from the audio program (a) based on the specified segments obtained in the feature description matching section 42. At this time, if the feature type specified by the search query is a thumbnail, feature values are not extracted from the audio program (a) but the thumbnail 1b is inputted into the feature extraction section 43.

The feature values or thumbnail 1b corresponding to the specified segments of the audio program (a) obtained in the feature extraction section 43 is fed into a feature presentation section 44 which plays and displays audio data corresponding to the feature values specified by user.

As can be seen, in this embodiment, using the feature description file 1a and/or the thumbnail 1b according to the present invention, audio data can be searched and browsed at various levels from the coarse level to the fine level. High-speed, efficient search and browsing can be achieved, accordingly.

Fig. 11 shows an alternative of the present invention. In this alternative, the contents of the clip description section 15, the stream description section 17, the object description section 19 and the event description section 21 are also added to the feature description file 1a.

As is obvious from the above description, according

from multiple pieces of audio video data are fed to feature description extraction sections 63 and 64, respectively. The feature description extraction sections 63 and 64 extract corresponding feature descriptions d1 and d2 from the feature description files b1 and b2 based on a certain feature type (c), respectively. Here, the feature type (c) to be extracted may be specified by a user's external input or feature descriptions may be described based on all feature types described in each feature description file. A feature description collection construction section 65 constructs a feature description collection (e) using multiple feature description files d1 and d2, and feeds the extracted feature description collection (e) to a feature description collection file generation section 66. The feature description collection file generation section 66 constructs a description as a feature description collection file using the description method according to the present invention, and generates a feature description collection file (f).

Fig. 14 shows a concrete example of the feature description collection (e) obtained from the present invention. In this example, the feature type (c) corresponds to a summary type for the individual audio video data a1 and a2, and examples for describing summaries based on a certain summary type (key event, home run) are shown. Summary descriptions are collected based on a targeted summary type from the audio video program collections 81,

82, (program 1, program 2,) and a summary collection 85 are constructed. For example, summary descriptions "home run" of the summary type are collected and the summary collection 85 consisting of 60th home run, 61st home run, 5 62nd home run, ... of a player named S. S is constructed.

As shown in Fig. 22, a summaries can be conventionally described only based on various summary types (key events, key objects and so on) for individual audio video programs (complete audio video programs). According to the present 10 invention, by contrast, summary descriptions can be collected from multiple audio video programs 81, 82, ... according to a specific summary type and the summary collection 85 can be thereby constructed and described.

Figs. 15A, 15B, 16A and 16B show feature description 15 collections which are described using a conventional feature description method and those according to the present invention. As shown in Fig. 15A, in a conventional feature description collection 91, audio video program identifiers 92a, 92b, ... referred to by each feature is described at 20 the highest level and feature types and contents as well as audio video data segments corresponding to the feature are described at the lower level. When the feature description collection 91 is browsed, the feature description collection file thus described is inputted into 25 and parsed by an audio video data browsing system. If, for example, "feature type 1" 93a, 93b, ... in the feature description collection are to be browsed, there is no means

for determining summaries based on the "feature type 1" 93a, 93b, ... are described in the programs represented by which identifiers 92a, 92b, Due to this, it is necessary to parse the feature description file 91 thoroughly from the beginning to the end. Further, if it is unclear in which range a reference program belonging to each feature type is valid and if many feature types exist, then it is sometimes difficult to specify the "feature type 1" 93a, 93b,

In a feature description collection 95 shown in Fig. 15B according to the present invention, by contrast, feature types and contents 93, 94, ... are described at the highest level and audio video program identifiers 92a, 92b, ... referred to by each feature based on the feature types 93, 94, ... and specified segments are described at a lower level under that of the feature types and contents. Accordingly, if a feature description collection based on a specific feature type and content, e.g., "feature type 1" 93 is to be browsed, it is enough to interpret only the highest level. If the highest level does not conform to the desired feature type and content, the elements are skipped until the next feature type 94. Once a desired feature description collection is searched, parsing can be finished at that point.

Further, since the reference programs 92a, 92b, ... are contained for each feature type 93, 94, ... , a program to be referred to can be easily specified. Further, while two "feature type 1" (93a, 93b) exist in the conventional feature

description collection 91, only one "feature type 1" exists in the feature description collection according to the present invention. It is, therefore, possible to avoid the overlapped description of the feature type 93 and to reduce
5 the size of the feature description collection file. Figs. 16A and 16B show the same contents as those of Figs. 15A and 15B in a table form, which description will not given herein.

Fig. 17 shows a construction and a processing flow if
10 the feature type (c) is "summary type (c)'" shown in Fig. 13. In this concrete example, summary description sections 71 and 72 describe the summaries of audio video programs a1' and a2', respectively. Summary description extraction sections 73 and 74 extract summary descriptions d1' and d2'
15 according to a certain summary type (key event, home run,) (c)' from summary description files b1' and b2' obtained by the summary description sections 71 and 72, respectively. A summary collection construction section 75 collects these summary descriptions d1' and d2' and constructs a summary
20 collection (e)'. Summary collection file generation section 76 generates a summary collection file (f)' using a summary collection description method according to the present invention.

Fig. 18A shows a feature description collection which
25 is described using the conventional feature description method as in the case of Fig. 15A. Fig. 18B shows a feature description collection which is described according to the

present invention as in the case of Fig. 15B.

In a summary collection file 101 according to the present invention, the summary type (c)' is set as "summary type: key event, content: home run" in Fig. 17, thereby
5 obtaining the first summary collection 102 from the summary collection construction section 75. Then, the summary type (c)' is set as "summary type: key event, content: two-base hit", thereby obtaining the second summary collection 103 from the summary collection construction section 75. The
10 summary collection file generation section 76 edits the first and second summary collections 102 and 103 into a summary collection file 101 and outputs the file 101. Through the above operations, the summary collection file 101 shown in Fig. 18B can be obtained.

15 Figs. 19A and 19B are illustrations for another embodiment according to the present invention. Fig. 19A shows a feature description collection described using the conventional feature description method. Fig. 19B shows a feature description collection described by the method
20 according to the present invention.

As shown in Fig. 19A, in the conventional feature description collection, program identifiers are described at the highest level and corresponding feature types and contents are described in parallel at the same level.
25 According to such a description method, it is difficult to extract a desired feature type by combining multiple feature types and contents.

In the feature description collection according to the present invention shown in Fig. 19B, by contrast, feature types and contents are described altogether and different feature types and contents are inserted in a nested structure, whereby it is possible to generate a feature description collection according to the different feature types or contents for the same feature type.

Figs. 20A and 20B are illustrations of the "feature type" shown in Fig. 19 is "summary type". Fig. 20A shows a feature description collection described using the conventional feature description method and Fig. 20B shows a feature description collection described according to the method of the present invention.

As shown in Fig. 20A, in the conventional summary collection description, program identifiers are described at the highest level and corresponding summary types and contents are described in parallel at the same level. According to such a description method, it is difficult to extract a desired summary description by combining multiple summary types and contents.

In the summary collection description according to the present invention shown in Fig. 20B, by contrast, summary types and contents are described altogether and different summary types and contents are inserted in a nested structure, whereby summaries can be described according to the different summary types or contents for the same summary type. For example, in the example shown in Fig. 20B, summaries are

described while nested "key event" 105 and "key objects" 106a and 106b.

Fig. 21 is a flow chart showing the outline of the operation of Fig. 17 in this embodiment. In a step S1, it
5 is judged whether or not nested structure is indicated to the summary collection construction section 75. If the judgment result of the step S1 is No, an operation for generating the summary collection file shown in Fig. 15B as already described above is carried out. If the judgment
10 result of the step S1 is Yes, a step S2 follows and a parent summary type (c)' is set. In a step S3, the summary description extraction sections 73 and 74 extract summary descriptions corresponding to the parent summary type (c)' from AV (audio video) programs 1 and 2, respectively. In
15 a step S4, a child summary type (c)' is set. In a step S5, the summary description extraction sections 73 and 74 extract summary descriptions corresponding to the child summary type (c)' from the AV programs 1 and 2, respectively. In a step
20 S6, summary types and contents are nested based on the extracted summary descriptions. In a step S7, it is judged whether or not all summary types have been set. If the judgment result of the step S7 is No, the step S2 follows and the procedures in the steps S2 through S6 are repeated. In this way, one or multiple summary collections with a nested
25 structure are formed. If the judgment result of the step S7 is Yes, a step S8 follows. In the step S8, the summary collection file generation section 76 generates a summary

collection file as shown in Fig. 20B.

With such a nested structure, it is possible to efficiently describe summaries based on multiple different summary types and contents, and to intelligently search and
5 browse audio video data.

As is evident from the above description given so far, according to the present invention, feature descriptions from multiple audio video programs are collected according to a specific feature type. Due to this, in case of
10 describing as a feature description collection, the feature descriptions can be represented efficiently and clearly. Further, it is possible to combine multiple feature types and to obtain a desired feature description from the feature description collection.

15